Proc. Natl. Acad. Sci. USA
Vol. 73, No. 5, pp. 1518–1522, May 1976
Biochemistry

Electron microscopy of gene regulation: The L-arabinose operon

(RNA nucleotidyltransferase/DNA fragments/transcription/positive control)

JAY HIRSH AND ROBERT SCHLEIF

Department of Biochemistry, Brandeis University, Waltham, Massachusetts 02154

Communicated by William P. Jencks, February 23, 1976

ABSTRACT **Using high magnification electron microscopy, we have observed protein complexes specifically bound to homogeneous DNA fragments containing the regulatory region of the *Escherichia coli* L-arabinose operon. Formation of detectable complexes on the DNA requires *araC* protein, catabolite gene activator protein, and RNA polymerase (nucleotidyltransferase). These protein–DNA complexes appear to be involved in regulation of the operon. Incubation of the proteins and DNA under inducing conditions leads to formation of a complex that appears to be capable of transcription, whereas incubation of the proteins and DNA under repressing conditions leads to formation of a different complex incapable of transcription. Using three overlapping DNA fragments, we located the sites of complex formation on the DNA. The complexes are formed on sequences 108 base pairs apart, an unexpectedly large distance.**

Normal induction of the *Escherichia coli* L-arabinose operon structural genes *araB*, *A*, and *D* requires the specific activator *araC* protein, and the general activator catabolite gene activator protein (CAP) (1–4). Most likely, both proteins assist initiation of transcription by RNA polymerase (nucleotidyltransferase) or act at a step that occurs soon after initiation (5). In addition to the inducing ability elicited by L-arabinose, *araC* protein possesses a repressing capability (6, 2, 3). D-Fucose, a close structural analogue of L-arabinose, converts *araC* protein into this form (7, 2). Through genetic and physiological studies a DNA region necessary for this repression, *araO*, has been located. It is transcriptionally ahead of the site necessary for normal induction of the operon, *araI* (6, 8, 9).

The regulatory region of DNA on which these induction and repression activities occur has been estimated by a combination of genetic and physical measurements to be about 300 base pairs long (9). This large size raises the possibility that regulation is accomplished not by proteins functioning together in a complex, but instead by regulatory proteins spaced along the DNA. For example, they could act sequentially as the RNA polymerase passes by.

We have initiated investigations of the mechanisms of arabinose operon induction and repression, using high resolution electron microscopy. This study has been made feasible by the development of electron microscopic techniques to visualize proteins specifically bound to short DNA fragments (refs. 10 and 11; J. Hirsh, and R. Schleif, manuscript submitted), by the isolation of short, homogeneous DNA fragments containing the *ara* regulatory region (12), and by the purification of active *araC* protein (D. Steffen, J. Greenblatt, and R. Schleif, manuscript submitted). We are able to visualize and localize two types of nonoverlapping complexes, each of which covers about 45 base pairs of DNA. The centers of the two regions where the two types of complexes form are separated by about 110 base pairs. Our evidence indicates that one complex is the induction complex, and the other may mediate repression. Genetic mapping has located the DNA region required for induction, *araI*, between the region required for repression, *araO*, and the transcribed genes, *araBAD* (6, 8, 9). The locations of the induction complex and putative repression complex agree with the genetic findings. Formation of both complexes requires the presence of *araC* protein, CAP protein, and RNA polymerase. These findings imply that the proteins act together in a complex to initiate transcription, and, most unexpectedly, that repression may require CAP and RNA polymerase in addition to *araC* protein.

Abbreviation: CAP, catabolite gene activator protein.

MATERIALS AND METHODS

Materials. The DNA fragments used were prepared by endonuclease S1 digestion of heteroduplex DNA prepared from λ*para* transducing phage (12, 9). The DNA molecules used in this study are diagrammed in Fig. 1. The 1120 base pair DNA carrying the *ara* control DNA and portions of the *araB* and *araC* genes was derived from the heteroduplex between DNA from λ*paraB114* and λ*paraC116*. The size of the DNA fragments was determined microscopically. A 203 base pair *lac* DNA fragment whose length is known from nucleotide sequencing was used as a standard (W. Gilbert and A. Maxam, personal communication). This DNA, mounted and viewed under our conditions, has an average interbase spacing of 0.270 ± 0.006 nm. The 500 base pair DNA with a deletion extending from the *araB* end was from the λ*paraB114*·λ*paraC135* heteroduplex, and the 740 base pair DNA with a deletion extending from the *araC* end was from the λ*paraC116*·λ*paraB209* heteroduplex. RNA polymerase was a gift of R. Burgess (13). CAP was a gift of J. Majors, prepared as described by Pastan *et al.* (14), with the exception that dithiothreitol was omitted from all buffers. The *araC* protein, obtained from D. Steffen, was purified through the DEAE-Sephadex step (Steffen, Greenblatt, and Schleif, manuscript submitted). This preparation is about 20% *araC* protein, as judged by sodium dodecyl sulfate polyacrylamide gel electrophoresis. Chemicals used were from Fisher and Sigma. The L-arabinose was found to contain contaminants that lead to heavy staining of grids. These were removed by passing the L-arabinose through a DEAE-cellulose column (DE-52, Whatman): 20 ml of 2 M L-arabinose were passed through a 3 ml column equilibrated in 10 mM Tris·HCl at pH 7.4.

Binding of Proteins to DNA. Unless otherwise indicated, the 20 μl reactions contained: 50 mM KCl; 10 mM $MgCl_2$; 10 mM Tris·HCl at pH 8.0; 1 mM EDTA; 1 mM dithiothreitol or dithioerythritol; 7.5% (vol/vol) glycerol; 0.2 mM cAMP (adenosine 3′:5′-cyclic monophosphate); 12.5 μg/ml of RNA polymerase; 20 μg/ml of CAP; 25 μg/ml of *araC* protein preparation; 0.5 μg/ml of DNA; and either 100 mM L-arabinose or 50 mM D-fucose. The proteins were incubated with DNA at 37° for 3 min. Poly(I) was then added to a final concentration of 0.75 μg/ml, and the incubation was continued for an additional 3 min. The reaction mixture was then passed at room temperature through a 0.3 ml agarose A-5m (Bio-Rad) column

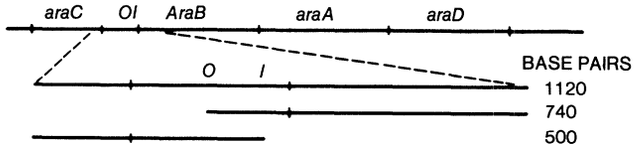

FIG. 1. Map of the L-arabinose operon, and the DNA fragments used in this work. The 1120 base pair DNA contains the *ara* regulatory region and portions of the *araC* and *araB* genes. The two other DNA molecules are a 740 base pair DNA lacking DNA sequences contained on the *araC* end of the 1120 DNA, and a 500 base pair DNA lacking the *araB* end of the 1120 DNA.

equilibrated in reaction mixture buffer. Approximately 2 min were required to pass samples through the column. Fractions, 20 μl each, from the agarose column were collected on a Saran Wrap (Dow Chemicals) drum, bound to isoamylamine-activated carbon grids (10), and stained with uranyl formate (11). Nucleoside triphosphates, 350 μM ATP, GTP; 75 μM CTP, UTP, were added to the appropriate incubation mixtures 1.5 min after the addition of poly(I). After incubating at 37° for 1.5 min, the reaction mixtures were passed through the agarose column and prepared for microscopy as described above.

Electron Microscopy. Grids were examined in a Philips 300 electron microscope. Micrographs were made at a film magnification of 30,000× on Kodak 5302 35 mm film; the contrast was enhanced by photographing on Kodak 5069 film, rephotographing on Pan-X film, and printing on Agfa grade 6 paper. Magnifications were calibrated in each roll of film by taking micrographs of paracrystalline tropomyosin (gift of P. Norton). Protein binding positions are given as the mean $\pm 2\sigma/\sqrt{n}$, where n is the number of molecules measured. This interval has a 0.95 probability of containing the true mean. Full details of the techniques will be published (J. Hirsh and R. Schleif, manuscript submitted).

RESULTS

Locations of the Bound Proteins. To study the mechanisms of induction and repression of the arabinose operon we have investigated protein complexes that form on the *ara* DNA. We sought to determine by electron microscopic methods the conditions necessary for formation of detectable complexes containing regulatory proteins or RNA polymerase on the DNA, and the locations of the complexes. Several species of DNA molecules were used to determine the locations of the complexes with respect to the *araC* and *araB* genes. Fig. 1 shows the L-arabinose operon of *E. coli* and the three different species of DNA molecules containing the *ara* regulatory region used for this determination. The longest species is 1120 base pairs and contains part of the *araC* gene, the regulatory region, and part of the *araB* gene. The two shorter species are subfragments of the 1120 base pair fragment as shown in Fig. 1.

RNA polymerase, CAP, *araC* protein, and the 1120 base pair DNA were incubated in buffer containing L-arabinose. The conditions used were similar to those used in demonstrating *araC*-dependent specific transcription of the arabinose operon (4). About one quarter of the DNA molecules were observed to contain a complex of bound protein(s). These complexes are about 12 nm diameter and are located 544 ± 8 nucleotides from an end of the DNA (Fig. 2a). Results obtained with the DNA having the *araC* gene end deleted (Fig. 2b) show that the complexes are located 544 base pairs from the *araC* gene end of the 1120 base pair DNA. Hereafter, these will be termed induction complexes.

araC protein, in the presence of D-fucose or the absence of L-arabinose, leads to repression of the transcription of the L-arabinose operon structural genes (7, 2). We therefore substituted D-fucose for L-arabinose while keeping all other components identical to those used for formation of induction complexes. Indeed, the induction complexes vanished and different complexes appeared. These were located with the 1120 base pair DNA fragment and the fragment with the *araB* end deleted. Fig. 2c and d shows the binding position histograms that locate the "repression" complexes 436 ± 8 base pairs from the *araC* gene end of the 1120 base pair DNA. The induction and the repression complexes each appear to form at unique sites on the DNA*. Omission of both D-fucose and L-arabinose was not attempted as purified *araC* protein is highly labile in the absence of these sugars.

Plate 1a and 1b shows micrographs of the 1120 base pair

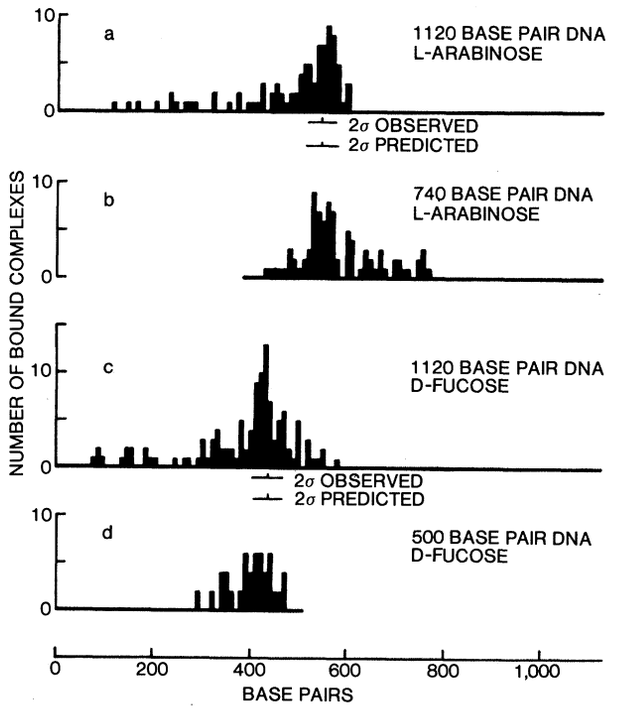

FIG. 2. Binding position histograms of protein complexes formed in the presence of *araC* protein, CAP, and RNA polymerase on the *ara* DNA fragments. All histograms have been normalized to show the frequency of protein complexes bound per 400 DNA molecules. The positions show the absolute distance from the approximate center of mass of the complex on the DNA to the nearest end of the DNA molecule, using 0.265 nm as the interbase spacing. The abscissa in each histogram shows the mean DNA length. The mean and observed standard deviations of the positions of specific binding were calculated considering molecules bound between 480 and 597 base pairs in a and 372–499 base pairs in c. The number of DNA molecules actually examined and the overall binding frequency are given after the description of each histogram. (a) *araC* protein, CAP, and RNA polymerase were incubated with the 1120 base pair DNA in buffer containing L-arabinose (388, 0.23). (b) As a, except the proteins were incubated with the 740 base pair DNA (482, 0.18). (c) *araC* protein, CAP, and RNA polymerase were incubated with the 1120 base pair DNA in buffer containing D-fucose (347, 0.33). (d) As c, except that the proteins were incubated with the 500 base pair DNA (231, 0.12).

* By our microscopy methods the standard deviation, σ, about the mean measured length, L, of a segment of DNA is $\sigma = 1.4 \times L^{1/2}$, where σ and L are in base pairs. The measured deviations shown in Fig. 2 closely agree with this relation. Numerical reconstructions show that the standard deviations would have been appreciably larger than those observed if the introduction or repression complex binding were not at unique sites, but instead were equally partitioned between two sites separated by 50 base pairs.

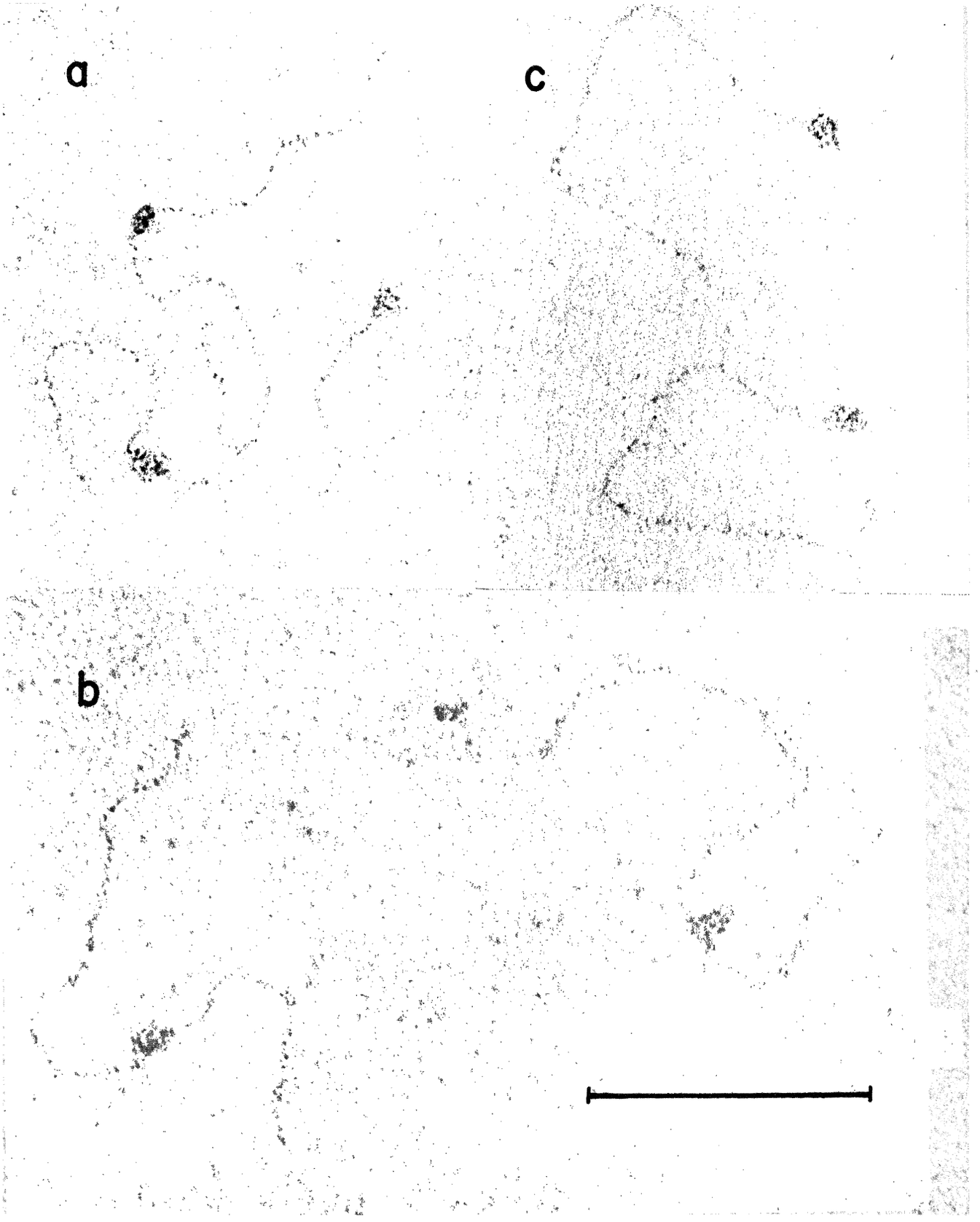

PLATE 1. Micrographs of *ara* regulatory protein–DNA complexes on the 1120 base pair DNA, and RNA polymerase bound to the phage λ promoter p_r. The bar represents 100 nm. (a) Protein–DNA complexes formed in the presence of *araC* protein, CAP, RNA polymerase and L-arabinose. One of the DNA molecules contains an RNA polymerase molecule bound at its end. This is observed on about 20% of the DNA molecules. (b) Protein–DNA complexes formed in the presence of *araC* protein, CAP, RNA polymerase, and D-fucose. (c) RNA polymerase molecules bound at the bacteriophage λ promoter p_r located 110 base pairs from the end of a 970 base pair DNA fragment.

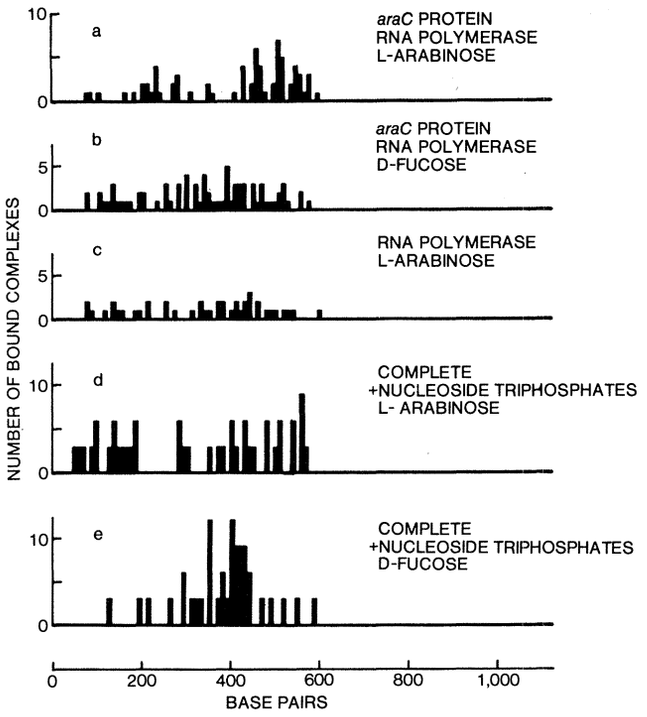

FIG. 3. Binding position histograms of proteins bound to the 1120 base pair DNA. (a) *araC* protein and RNA polymerase were incubated with DNA in buffer containing L-arabinose (342, 0.21). (b) *araC* protein and RNA polymerase were incubated with DNA in buffer containing D-fucose (507, 0.15). (c) RNA polymerase was incubated with DNA in buffer containing L-arabinose (424, 0.10). (d) *araC* protein, CAP protein, and RNA polymerase were incubated with DNA in buffer containing L-arabinose, then nucleoside triphosphates were added (137, 0.28). The higher background in this experiment may represent RNA polymerase molecules caught in the act of transcribing. (e) as d, except that buffer contained D-fucose instead of L-arabinose (145, 0.25).

molecules containing the complexes formed in the presence of L-arabinose or D-fucose†. Plate 1c, at the same magnification, is a micrograph of RNA polymerase bound to the promoter p_r of bacteriophage λ‡.

What Is Needed for the Proteins to Bind. All three proteins are required to form either the induction or repression complexes. Omission of any of the proteins or any combination of two proteins (or all three proteins) greatly reduced the frequency and specificity of complex formation on the DNA. Fig. 3a, b, and c shows portions of these data. Whenever RNA polymerase was omitted, none of the 12 nm complexes was observed. Despite an intensive search for *araC* protein or CAP, which may bind to specific sequences (15, 16) in the absence of RNA polymerase, little binding and no specific binding of particles with diameter 4.0 nm or greater was observed. From

† In the molecules displayed in Plate 1, DNA passes through the edge of the induction complexes and through the center of the representation complexes. These predilections were not maintained in larger samples, and examination of several hundred complexes failed to reveal any regular differences between the induction and repression complexes.

‡ Approximately 20% of the RNA polymerase molecules we have observed on the p_r promoter have an "arrowhead" shape, all of which are pointed in the direction of subsequent transcription. The remaining 80% are nondescript "blobs." One of the RNA polymerase molecules bound to λp_r in Plate 1 possesses the "arrowhead" morphology. Clear "arrowheads" appear less often in the *ara* induction or repression complexes than they do on the lambda p_r promoter.

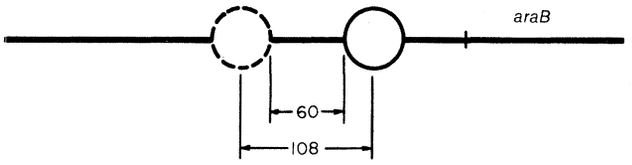

FIG. 4. A scale drawing of the protein complexes bound to the *ara* regulatory region DNA. This DNA represents the central 500 base pairs of the 1120 base pair DNA. The solid circle represents the induction complex, and the broken circle represents the repression complex. The distances are given in base pairs. The exact position of the beginning of the *araB* gene is not known.

the appearance of the 12 nm RNA polymerase on the 1.5 nm diameter DNA, we estimate that we would have observed particles of diameter 4 nm or greater. In solution *araC* protein and CAP sediment as though their molecular masses were 60,000 and 45,000 daltons, respectively (D. Steffen, J. Greenblatt, and R. Schleif, manuscript submitted; ref 17), and if the proteins are roughly spherical when bound to DNA, each would have a diameter of about 5 nm. Most likely then, neither *araC* protein nor CAP binds to the *ara* regulatory region with sufficient tenacity to be observed by our techniques, or each stains less well than any other proteins we have studied.

The Induction Complexes Can Transcribe. We have previously found that these electron microscopy techniques can be used to show that RNA polymerase molecules bound to the bacteriophage lambda promoter p_r are transcriptionally active (J. Hirsh, and R. Schleif, manuscript submitted). This was done by adding nucleoside triphosphates to an incubation mixture containing RNA polymerase bound at the promoter p_r. After a brief 37° incubation, the mixture was shifted to 20°, a temperature where transcription by a bound polymerase proceeds nearly normally, but where the binding of RNA polymerase to a vacant promoter is very slow. The DNA was then mounted and stained as usual. The DNA lacked polymerase at the promoter if the triphosphates were added but contained the polymerase if rifampin was added before the triphosphates. Both properties are in accord with present knowledge of transcription (18). The results of an analogous transcription experiment performed with the *ara* system are shown in Fig. 3d and e. Nucleoside triphosphates removed the induction complexes, while they had no effect on the repression complexes. Thus, while the repression complexes are likely to contain an RNA polymerase molecule, they are incapable of transcription under these conditions.

DISCUSSION

Utilizing high resolution electron microscopy, we have observed and located two types of protein–DNA complexes on DNA containing the regulatory sequences of the *E. coli* L-arabinose operon. L-Arabinose, *araC* protein, CAP, and RNA polymerase are required to form the complex which appears to be an intermediate in the normal induction process. Therefore, *araC* protein and CAP must exert a control over the initiation of transcription rather than acting exclusively after transcription has begun. Substituting D-fucose for L-arabinose promotes instead the formation of a "repression" complex about 110 nucleotides ahead of the induction complexes. This region has previously been shown to be required for repression (6–8). Fig. 4 shows a drawing to scale of the complexes' binding positions on the *ara* regulatory region DNA. If the "repression" complex indeed represses, it must function differently from the *lac* and λ phage repressors that bind to DNA directly overlapping the RNA polymerase tight binding sites (refs. 19–21; J. Gralla,

manuscript in preparation). Possibly, the proteins forming the induction complex bind transiently to a site upstream from the location at which we observe the induction complex. The "repression" complex could act by blocking access to this preliminary site or by preventing the translocation to the induction site.

We did not observe RNA polymerase binding to the promoter for synthesis of *araC* protein. This implies that binding to the *araC* promoter is weak, slow, or requires additional proteins or small molecules. The *araC* promoter most likely lies toward *araC* from the induction site. In view of the transcriptional inactivity of our "repression" complex and its formation only in the presence of D-fucose, we consider it unlikely but not impossible that the complex is instead involved with synthesis of *araC* protein.

The transcriptional ability of the L-arabinose-dependent complexes, their location, and their requirements for formation strongly link them to induction. However, the evidence that the D-fucose-dependent complexes are related to repression is less solid. It rests upon the complexes' D-fucose and *araC* requirement, their location, and their inability to initiate transcription when supplied with nucleoside triphosphates. If these D-fucose-dependent complexes are truly repression complexes, then several interesting questions can be raised. Why are CAP and RNA polymerase required for their formation? How does the complex accomplish repression from its location 108 nucleotides on the upstream side of the induction complex? What is the explanation for the finding that *araC* protein alone binds to *araO* DNA as observed by retention on Millipore filters (15)?

This work was supported in part by the National Institutes of Health (Research Grant GM18277, Career Development Award KO4-GM-38797, and Training Grant GM00212) and the Dreyfus Teacher-Scholar Award. We thank R. Burgess, S. Kaplan, J. T. Lis, J. Majors, P. Norton, and D. Steffen for gifts of materials used in this study. This is Publication no. 1065 from the Department of Biochemistry, Brandeis University.

1. Sheppard, D. E. & Englesberg, E. (1967) *J. Mol. Biol.* **25,** 443–454.
2. Greenblatt, J. & Schleif, R. (1971) *Nature New Biol.* **233,** 166–170.
3. Wilcox, G., Meuris, P., Bass, R. & Englesberg, E. (1974) *J. Biol. Chem.* **249,** 2946–2952.
4. Lee, N., Wilcox, G., Gielow, W., Arnold, J., Cleary, P. & Englesberg, E. (1974) *Proc. Natl. Acad. Sci. USA* **71,** 634–638.
5. Hirsh, J. & Schleif, R. (1973) *J. Mol. Biol.* **80,** 433–444.
6. Englesberg, E., Squires, C. & Meronk, F. (1969) *Proc. Natl. Acad. Sci. USA* **62,** 1100–1107.
7. Englesberg, E., Irr, J., Power, J. & Lee, N. (1965) *J. Bacteriol.* **90,** 946–956.
8. Kessler, D. P. & Englesberg, E. (1969) *J. Bacteriol.* **98,** 1159–1169.
9. Schleif, R. & Lis, J. T. (1975) *J. Mol. Biol.* **95,** 417–431.
10. Dubochet, J., Ducommun, M., Zollinger, M. & Kellenberger, E. (1971) *J. Ultrastruct. Res.* **35,** 147–167.
11. Brack, C. (1973) *Experientia* **28,** 768.
12. Lis, J. T. & Schleif, R. (1975) *J. Mol. Biol.* **95,** 409–416.
13. Burgess, R. & Jendrisak, J. (1975) *Biochemistry* **14,** 4634–4638.
14. Pastan, I., Gallo, M. & Anderson, W. (1974) in *Methods in Enzymology,* eds. Hardman, J. & O'Malley, B. (Academic Press, New York), Vol. 38, pp. 367–376.
15. Wilcox, G., Clemetson, K. J., Cleary, P. & Englesberg, E. (1974) *J. Mol. Biol.* **85,** 589–602.
16. Majors, J. (1975) *Nature* **256,** 672–674.
17. Anderson, W. B., Schneider, A. B., Emmer, M., Perlman, R. & Pastan, I. (1971) *J. Biol. Chem.* **246,** 5929–5937.
18. Hinkle, D. C., Mangel, W. F. & Chamberlin, M. J. (1972) *J. Mol. Biol.* **70,** 209–220.
19. Gilbert, W. & Maxam, A. (1973) *Proc. Natl. Acad. Sci. USA* **70,** 3581–3584.
20. Maniatis, T., Ptashne, M., Backman, K., Kleid, D., Flashman, S., Jeffrey, A. & Maurer, R. (1975) *Cell* **5,** 109–113.
21. Walz, A. & Pirrotta, V. (1975) *Nature* **254,** 118–121.